

Causal Parrots: Large Language Models May Talk Causality But Are Not Causal

Year: 2023 | Citations: 154 | Authors: M. Zecevic, Moritz Willig, D. Dhami

Abstract

Some argue scale is all what is needed to achieve AI, covering even causal models. We make it clear that large language models (LLMs) cannot be causal and give reason onto why sometimes we might feel otherwise. To this end, we define and exemplify a new subgroup of Structural Causal Model (SCM) that we call meta SCM which encode causal facts about other SCM within their variables. We conjecture that in the cases where LLM succeed in doing causal inference, underlying was a respective meta SCM that exposed correlations between causal facts in natural language on whose data the LLM was ultimately trained. If our hypothesis holds true, then this would imply that LLMs are like parrots in that they simply recite the causal knowledge embedded in the data. Our empirical analysis provides favoring evidence that current LLMs are even weak 'causal parrots.'